

CommonGen: A Constrained Text Generation Challenge for Generative Commonsense Reasoning

Year: 2020 | Citations: 389 | Authors: Bill Yuchen Lin, Ming Shen, Wangchunshu Zhou, Pei Zhou, Chandra Bhagavatula

Abstract

Recently, large-scale pre-trained language models have demonstrated impressive performance on several commonsense-reasoning benchmark datasets. However, building machines with commonsense to compose realistically plausible sentences remains challenging. In this paper, we present a constrained text generation task, CommonGen associated with a benchmark dataset, to explicitly test machines for the ability of generative commonsense reasoning. Given a set of common concepts (e.g., dog, frisbee, catch, throw); the task is to generate a coherent sentence describing an everyday scenario using these concepts (e.g., “a man throws a frisbee and his dog catches it”). The CommonGen task is challenging because it inherently requires 1) relational reasoning with background commonsense knowledge and 2) compositional generalization ability to work on unseen concept combinations. Our dataset, constructed through a combination of crowdsourced and existing caption corpora, consists of 77k commonsense descriptions over 35k unique concept-sets. Experiments show that there is a large gap between state-of-the-art text generation models (e.g., T5) and human performance (31.6% v.s. 63.5% in SPICE metric). Furthermore, we demonstrate that the learned generative commonsense reasoning capability can be transferred to improve downstream tasks such as CommonsenseQA (76.9% to 78.4 in dev accuracy) by generating additional context.